Regression Models for Lifetime Given the Usage History

Thierry Duchesne
Département de mathématiques et de statistique
Pavillon Alexandre-Vachon
Université Laval
Québec, QC, G1K 7P4
Canada
duchesne@mat.ulaval.ca

Abstract

In several reliability applications, a model for the distribution of lifetime given a realized (or forecast) usage history is necessary. Such examples include the choice of a time scale, the calculation of warranty prices or the extrapolation of an accelerated life test. There is a vast array of approaches proposed to model the relationship between lifetime and usage in the literature, be it through hazard function or internal wear modelling, or through time scale changes. This talk surveys some of these methods and outlines some of their main features, such as related inference procedures, probabilistic interpretation, applicability to certain problems, and so on.

1 Introduction

Let T be the random variable of time to failure of an item and let $\boldsymbol{\theta}_t = \{\theta(u), 0 \le u \le t\}$ represent the usage history up to time t for this item, where $\theta(u)$ represents the usage rate of the item at time u. Define the cumulative usage at time t as $y(t) = \int_0^t \theta(u)du$. For example, the item could be a car, T could be its failure time, $\theta(t)$ its speed at time t, $y(t)$ its cumulative mileage at time t and $\boldsymbol{\theta}_t$ its mileage accumulation history from time 0 up to time t.

Suppose that for a fleet of n independent items, we dispose of the following information: $(t_i, \delta_i, \{\theta_i(u), 0 \le u \le t_i\})$, $i = 1, \ldots, n$, where t_i is a failure or right-censoring time, δ_i is an indicator that t_i is a failure time and $\theta_i(u)$ is the usage rate of item i at time u. Our purpose in this paper is to use such data to infer about the distribution of lifetime, T, given the usage history, $\boldsymbol{\theta}_t$. In most situations, a regression model for the effect of usage on lifetime is postulated. Most regression models specify certain aspects of the relationship between lifetime and usage and leave unspecified some parts of this relationship (finite or infinite dimensional parameters), which are to be estimated from data. In this paper we consider several approaches to this regression problem and outline some of their properties.

2 Most common modelling approaches

Some approaches to modelling the effect of usage on lifetime are well known and have been well documented. Let $\boldsymbol{\Theta}_t$ denote the space of all possible usage accumulation histories between times 0 and t.

2.1 Proportional hazards model

Perhaps the best known model in lifetime regression, it assumes that the effect of usage accumulation on lifetime acts multiplicatively on the hazard of failure at time t:

$$h(t|\boldsymbol{\theta}_t) = h_0(t)\psi(\boldsymbol{\theta}_t; \boldsymbol{\beta}), \tag{1}$$

where $h_0(t)$ is a baseline hazard rate, usually left arbitrary, and $\psi : \boldsymbol{\Theta}_t \to I\!R$ maps the usage history up to time t to its multiplicative effect on the hazard and usually depends on a finite dimensional parameter $\boldsymbol{\beta}$. Banjevic et al. (2001) use this model with $\psi(\boldsymbol{\theta}_t; \boldsymbol{\beta}) = \exp\{\beta\theta(t)\}$ to optimize condition-based maintenance.

2.2 Additive hazards model

Frequently used in reliability to model the hazard of an event at time t given a realization of a stochastic process:

$$h(t|\boldsymbol{\theta}_t) = h_0(t) + \psi(\boldsymbol{\theta}_t; \boldsymbol{\beta}),$$

where $\psi(\cdot)$ is as in (1). Singpurwalla and Wilson (1993, 1998) use this model with $\psi(\boldsymbol{\theta}_t; \boldsymbol{\beta}) = \beta y(t)$ for two-dimensional warranty calculations. Cox (1999) derives some properties of this model.

2.3 Accelerated failure time model

Though its version with time-fixed covariates is better known, this model is also well defined for time-varying covariates:

$$P[T > t|\boldsymbol{\theta}_t] = G\left(\int_0^t \psi(\boldsymbol{\theta}_u; \boldsymbol{\beta}) du\right), \tag{2}$$

where $G(\cdot)$ is a survivor function. Properties and semiparametric inference for this model are discussed by Robins and Tsiatis (1992). Lawless et al. (1995) use the time-fixed covariate version to analyze two-dimensional automobile warranty data.

3 Other modelling approaches

The models of this section are perhaps not as well known as the models from Section 2, but they are nonetheless quite interesting and deserve, in the author's opinion, further consideration.

3.1 Ideal time scales

Duchesne and Lawless (2000) define an ideal time scale (ITS) as a functional $\phi(t, \boldsymbol{\theta}_t)$ such that

$$P[T > t|\boldsymbol{\theta}_t] = G[\phi(t, \boldsymbol{\theta}_t)],$$

where $G(\cdot)$ is a survivor function that does not depend on $\boldsymbol{\theta}_t$. Other authors have referred to such time transformations as *intrinsic scale* (Çinlar and Ozekici, 1987), *load invariant scale* (Kordonsky and Gertsbakh, 1997) or *virtual age* (Finkelstein, 1999) and is in close correspondence with the concepts of *transfer functional* and *resource* of Bagdonavičius and Nikulin (1997).

3.2 General models of Bagdonavičius and Nikulin (1997)

Bagdonavičius and Nikulin (1997) propose various classes of general models that include the proportional hazards, additive hazards, accelerated failure time and other time scale change models. They obtain their classes of models by defining a *transfer functional*, $f_{\boldsymbol{\theta}_t}(t)$, which is closely related to an ITS, via differential equations. They derive inference methods and outline goodness-of-fit procedures for such models.

3.3 Stochastic internal wear and failure rate

Singpurwalla (1995) models the reliability of items by assuming that their internal wear (degradation, system state, health level, etc.) and/or the hazard of failure are stochastic processes influenced by the environment where the items live. If we let $\{X(t), t \geq 0\}$ represent the value of the internal wear of an item and X^* be the (perhaps random) *failure threshold*, then time to failure can be defined as $T = \inf\{t : X(t) \geq X^*\}$. Though there is a vast literature on models for T based on models for $\{X(t)\}$ and on the modelling of internal wear or degradation as a function of covariates (Meeker and Escobar, 1998, chapter 21), few articles seem to unify both concepts to model reliability as a function of the usage history.

Bagdonavičius and Nikulin (2001) do unify degradation and regression models for this purpose. They let the time index of the process $\{X(t)\}$ depend on $\boldsymbol{\theta}_t$. More precisely, they consider a gamma diffusion model for $\{X(t)\}$ under a "baseline" usage $\boldsymbol{\theta}^*$, and the effect of a different usage $\boldsymbol{\theta}$ is modelled by altering the value of the time index t by using $t^* = \int_0^t \exp\{\beta\theta(u)\}du$ instead. This can be viewed as an application of the accelerated failure time model (2) to model the effect of usage on the degradation of items. Note that if we consider usage as a stochastic process, then this is an interesting example of a *subordinated stochastic process*. Lee and Whitmore (1993) discuss properties of these processes of the form $X(T(t))$, in the case where $\{X(t)\}$ is a continuous and stationary Markov process and $\{T(t)\}$ is a process with non-negative and independent increments. Hougaard et al. (1997) combine a Poisson and a Hougaard process in this manner to analyze epileptic seizure data; random time can be viewed as within patient variability and thus an approach based on subordinated processes can be useful for situations where overdispersion is present. And clearly, from Bagdonavičius and Nikulin (2001), subordinated processes have the potential to be useful in developing models and methods for regression of lifetime given usage.

Singpurwalla (1995), Cox (1999) and Bagdonavičius and Nikulin (2001) model the effect of covariates such as the usage rate on failure not only through the value of the internal wear process $\{X(t)\}$, but also by modelling the effect of $\boldsymbol{\theta}_t$ on the distribution of the failure threshold, X^*, or equivalently on the hazard that a traumatic event that may kill the item occurs.

4 Collapsible models

This class of models is little known but can potentially be useful in modelling the distribution of lifetime given a usage history. A regression model is said to be *collapsible* when

$$P[T > t|\boldsymbol{\theta}_t] = G[\phi(t, y(t); \boldsymbol{\beta})],$$

where $\phi(\cdot, \cdot; \boldsymbol{\beta}) : \mathbb{R}^2 \to \mathbb{R}$ is a positive map and $G(\cdot)$ is a survivor function. Oakes (1995) introduced these models and used the model with a linear scale $\phi(t, y(t); \beta) = t + \beta y(t)$ to model the lifetime of miners given their history of exposure to asbestos dust.

4.1 Statistical inference

Oakes (1995), Kordonsky and Gertsbakh (1997), Duchesne (2000) and Duchesne and Lawless (2000) discuss maximum likelihood based inference methods for this model when the survivor function $G(\cdot)$ is specified parametrically. In the semiparametric case (i.e., $\phi(\cdot, \cdot; \boldsymbol{\beta})$ parametric but $G(\cdot)$ arbitrary), Kordonsky and Gertsbakh (1997) derive inference methods based on the coefficient of variability of the age in the ITS while Duchesne and Lawless (2002) propose a method based on ranks (counting processes). In the nonparametric case (i.e., both $\phi(\cdot, \cdot)$ and $G(\cdot)$ arbitrary), Duchesne (2000) proposes an ad hoc method to estimate the level curves $C_t = \{(x, y) : \phi(x, y)] = t\}$ (see Figure 1). The points on such an age curve correspond to the same quantile of $P[T > t|\boldsymbol{\theta}_t]$.

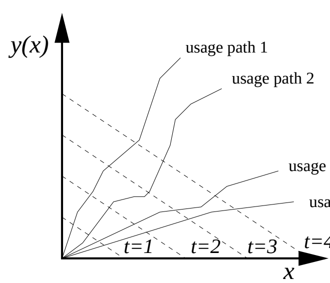

Figure 1: Example of age curves from a collapsible model with a linear ideal time scale $\phi(x, y(x)) = x + \beta y(x)$.

4.2 Stochastic justification of the model

Duchesne and Rosenthal (2003) consider the dynamic environment setup of Singpurwalla (1995), Cox (1999) and Bagdonavičius and Nikulin (2001). Assuming that the internal wear follows a diffusion process whose drift depends on the usage history, they obtain collapsible models under certain conditions. They also consider cases with traumatic events whose rate of occurrence depends on usage.

4.3 Potential for applications

Collapsible models have been used to model the reliability of several items, such as aircraft or steel specimens (Kordonsky and Gertsbakh, 1997). They are also useful for preventive maintenance decisions in two dimensions, as investigated by Frickenstein and Whitaker (2003). Moreover, the age curves of Figure 1 suggest a shape for what could be viewed as "fair" warranty regions. Note that if lifetime given usage followed a collapsible model in a case where the warranty region is parallel to the age curves, then warranty cost calculations and inference from warranty data could be much simplified.

Acknowledgements

This work was supported by individual operating grants from the Natural Sciences and Engineering Research Council of Canada and of the Fonds québécois de la recherche sur la nature et les technologies.

References

Bagdonavičius, V.B. and Nikulin, M.S. (1997). Transfer functionals and semiparametric regression models. *Biometrika 84*, 365–378.

Bagdonavičius, V.B. and Nikulin, M.S. (2001). Estimation in degradation models with explanatory variables. *Lifetime Data Analysis 7*, 85–103.

Banjevic, D., Jardine, A.K.S., Makis, V. and Ennis, M. (2001). A control-limit policy and software for condition-based maintenance optimization. *INFOR 39*, 32–50.

Çinlar, E. and Ozekici, S. (1987). Reliability of complex devices in random environments. *Probability in the Engineering and Informational Sciences 1*, 97–115.

Cox, D.R. (1999). Some remarks on failure-times, surrogate markers, degradation, wear and the quality of life. *Lifetime Data Analysis 5*, 307–314.

Duchesne, T. (2000). Semiparametric methods of time scale selection. In N. Limnios and M. Nikulin (Eds), *Recent Advances in Reliability Theory: Methodology, Practice and Inference*, pp. 279–289. Birkhauser: Boston.

Duchesne, T. and Lawless, J. (2000). Alternative time scales and failure time models. *Lifetime Data Analysis 6*, 157–179.

Duchesne, T. and Lawless, J. (2002). Semiparametric inference methods for general time scale models. *Lifetime Data Analysis 8*, 263-276.

Duchesne, T. and Rosenthal, J.S. (2003). On the collapsibility of lifetime regression models. *Advances in Applied Probability 35*, 755–772.

Finkelstein, M.S. (1999). Wearing-out of components in a variable environment. *Reliability Engineering and System Safety 66*, 25–242.

Frickenstein S.G. and Whitaker L.R. (2003). Age replacement policies in two time scales. *Naval Research Logistics 50*, 592–613.

Houugaard, P., Lee, M.-L. T. and Whitmore, G.A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics 53*, 1225–1238.

Kordonsky, K.B. and Gertsbakh, I. (1997). Multiple time scales and the lifetime coefficient of variation: engineering applications. *Lifetime Data Analysis 3*, 139–156.

Lawless, J.F., Hu, J. and Cao, J. (1995). Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis 1*, 227–240.

Lee, M.-L. T. and Whitmore, G.A. (1993). Stochastic processes directed by randomized time. *Journal of Applied Probability 30*, 302–314.

Meeker, W.Q. and Escobar, L.A. (1998). *Statistical Methods for Reliability Data*. New York, Wiley.

Oakes, D. (1995). Multiple time scales in survival analysis. *Lifetime Data Analysis 1*, 7–18.

Robins, J. and Tsiatis, A.A. (1992). Semiparametric estimation of an accelerated failure time model with time-varying covariates. *Biometrika 79*, 311–319.

Singpurwalla, N.D. (1995). Survival in dynamic environments. *Statistical Science 10*, 86–103.

Singpurwalla, N.D. and Wilson, S.P. (1993). The warranty problem: its statistical and game theoretic aspects. *SIAM Review 35*, 17–42.

Singpurwalla, N.D. and Wilson, S.P. (1998). Failure models indexed by two scales. *Advances in Applied Probability 30*, 1058–1072.